Exhibit B 09/843,159

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 262, No. 33, Issue of November 25, pp. 15990-15997, 1987
Printed in U.S.A.

Primary Structure of Human Poly(ADP-ribose) Synthetase as Deduced from cDNA Sequence*

(Received for publication, May 18, 1987)

Tomohiro Kurosaki‡, Hiroshi Ushiro‡, Yasuhiro Mitsuuchi‡, Shigetaka Suzuki‡, Michiko Matsuda‡, Yoshiko Matsuda§, Nobuhiko Katunuma§, Kenji Kangawa¶, Hisayuki Matsuo¶, Tadaaki Hirose‖, Seiichi Inayama‖, and Yutaka Shizuta‡

From the ‡Department of Medical Chemistry, Kochi Medical School, Nankoku, Kochi 781-51, Japan, the §Department of Enzyme Chemistry, Institute for Enzyme Research, School of Medicine, the University of Tokushima, Tokushima 770, Japan, the ¶Department of Biochemistry, Miyazaki Medical College, Miyazaki 889-16, Japan, and the ‖Pharmaceutical Institute, Keio University School of Medicine, Tokyo 160, Japan

Human poly(ADP-ribose) synthetase consists of three proteolytically separable domains, the first for binding of DNA, the second for automodification, and the third for binding of the substrate, NAD (Ushiro, H., Yokoyama, Y., and Shizuta, Y. (1987) *J. Biol. Chem.* 262, 2352–2357). We have isolated and sequenced cDNA clones for the enzyme using synthesized oligodeoxyribonucleotide probes based on the partial amino acid sequence of the protein. The open reading frame determined encodes a protein of 1,013 amino acid residues with a molecular weight of 113,203. The deduced amino acid sequence is consistent with the partial amino acid sequences of tryptic or α-chymotryptic peptides and the total amino acid composition of the purified enzyme. The native enzyme is relatively hydrophilic as judged from the hydrophilicity profile of the total amino acid sequence. The net charge of the NAD binding domain is neutral but the DNA binding domain and the automodification domain are considerably rich in lysine residue and quite basic. The DNA binding domain involves a homologous repeat in the sequence and exhibits a sequence homology with localized regions of transforming proteins such as c-*fos* and v-*fos*. Furthermore, this domain contains a unique sequence element which resembles the essential peptide sequences for nuclear location of SV40 and polyoma virus large T antigens. These facts suggest the possibility that the physiological function of poly(ADP-ribose) synthetase lies in its ability to bind to DNA and to control transformation of living eukaryotic cells like the cases of those oncogene products.

Poly(ADP-ribose) synthetase, an enzyme localized in the nucleus of eukaryotic cells, catalyzes the polymerization of the ADP-ribose moiety of NAD to form a bipolymer, poly(ADP-ribose), which is covalently bound to various nuclear proteins (1, 2). A unique feature of this enzyme is that it requires DNA for catalytic activity and that it is subjected to automodification during the reaction (3, 4).

Recently, the enzyme has been purified to homogeneity from various sources and extensively characterized (5–8). It has been demonstrated that the enzyme consists of three proteolytically separable domains, the first for binding of DNA, the second for automodification, and the third for binding of the substrate, NAD (7–12). Whereas the physiological function of this enzyme is not as yet fully understood, several lines of evidence suggest that it may be involved in many biologically important processes such as DNA repair, DNA replication, RNA synthesis, and cell differentiation (See Ref. 2). Nevertheless, how poly(ADP-ribosyl)ation participates in these important biological mechanisms and how the gene for poly(ADP-ribose) synthetase is regulated in eukaryotic cells remain to be elucidated. In order to provide an initial molecular genetic approach to investigate the physiological role of the enzyme in living cells, we have isolated cDNA clones representing most of a 4.9-kilobase mRNA for poly(ADP-ribose) synthetase in human placenta.

In this paper, we report the isolation of cDNA clones for the mRNA and present the nucleotide sequence of the cloned cDNAs which allows us to predict the complete amino acid sequence of this polypeptide. Structural characteristics of the three functional domains as deduced from cDNA sequence as well as the homology of the predicted amino acid sequence of each domain with those of other proteins are also discussed.

EXPERIMENTAL PROCEDURES[1]

RESULTS

cDNA and Protein Sequences—Fig. 2 shows the restriction map and the sequence strategy for the cloned cDNAs. The nucleotide sequences were determined on both strands of the cDNAs for all but 365 residues corresponding to the 3′ end of the mRNA; for this region, sequence determination on both strands was technically difficult, but the sequence data were reliable. Fig. 3 shows the 3792 nucleotide sequence (excluding the poly(dA) tract) of the cDNA encoding human poly(ADP-ribose) synthetase, determined using clones pPARS1, pPARS11, pPARS21, pPARS31, pPARS32, pPARS41, and

* This work was supported in part by grants from Arima Memorial Foundation for Medical Research, grants-in-aid for Cancer Research from Ohtsuka Pharmaceutical Industries Co., and grants-in-aid for Science and Cancer Research from the Ministry of Education, Science, and Culture in Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03473.

[1] Portions of this paper (including "Experimental Procedures," part of "Results," Figs. 1, 3, 6, and 7, and Tables I and II) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 87M-1665, cite the authors, and include a check or money order for $5.60 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

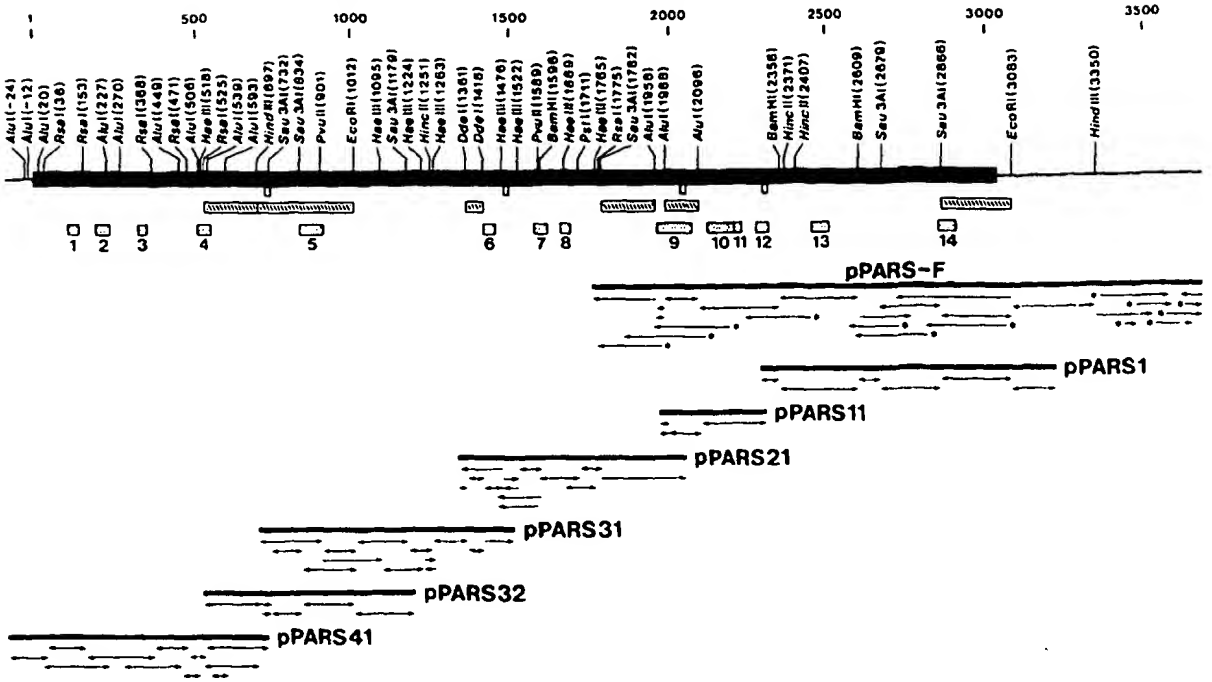

FIG. 2. **Strategy for sequencing cloned cDNA encoding human poly(ADP-ribose) synthetase.** The restriction map displays only relevant restriction endonuclease sites, identified by numbers indicating the 5'-terminal nucleotide generated by cleavage (for the nucleotide numbers, see Fig. 3). The poly(dA)·poly(dT) tract and the poly(dG)·poly(dC) tails are not included in the restriction map. The protein-coding region is indicated by a *closed box*, the sequence used for specific priming of reverse transcription by a *small open box*, and the sequence used as hybridization probes for selecting clones by a *hatched box*. The coding region verified by the amino acid sequence analysis using 14 different peptides of human placental poly(ADP-ribose) synthetase is shown by a *dotted box*. The direction and extent of sequence determinations are indicated by *horizontal arrows* under each clone used. The *star symbols* at the end of *arrows* denote that sequencing was performed using deletion mutants.

pPARS-F. No sequence difference was observed among the overlapped cDNA sequences determined with the individual clones. All of the 14 partial amino acid sequences determined by peptide analysis were found to be encoded by the cDNA sequence in the same reading frame (Figs. 1 and 2 and Table II).

The translational initiation site is assigned to be the methionine codon composed of nucleotide residues 1–3 because this is the first ATG triplet that appears downstream of the nonsense codon TAG (residue −9 to −7) found in the frame. This assignment is supported by the fact that the nucleotide sequence around the ATG triplet agrees well with the favored sequence that flanks functional initiation codons of $^{A}_{G}XXAUGG$, where X can be any nucleotide (30). The codon specifying the tryptophan residue at position 1013 is followed by the translational termination codon TAA. The length of the 3'-noncoding region of the cDNA excluding the poly(dA) tract was estimated to be 658 residues by sequence analyses using clones pPARS1 and pPARS-F.

From the cDNA sequence, we concluded that poly(ADP-ribose) synthetase from human placenta consists of 1,013 amino acid residues. The molecular weight of the protein was calculated to be 113,203. This value is in good agreement with that reported for the molecular weight of the enzyme from human placenta (8). Furthermore, the amino acid composition of the enzyme as deduced from the cDNA sequence reasonably agrees with that experimentally determined (Table III).

Northern Blot Analysis of the mRNA—To determine the size of the mRNA encoding poly(ADP-ribose) synthetase, a series of Northern hybridization experiments were performed using poly(A)$^+$ RNA from human placenta and restriction fragments of the cDNA inserts of clones pPARS1, pPARS21, pPARS31, and pPARS41. Only one size species of mRNA was found in all cases. The size of the mRNA for the enzyme was estimated to be 4.9 kilobase, based on its electrophoretic mobility relative to known standards (Fig. 4).

Characteristics of the Three Domains—Judging from the amino acid sequence of the N terminus of the 54-kDa fragment as determined by peptide analysis, we concluded that the NAD binding domain consists of 489 residues starting from residue 525 to the C-terminal tryptophan of the enzyme protein. The molecular weight of 54,881 calculated from the predicted amino acid sequence (Table III) coincides well with the experimentally determined value of 54,000 for the molecular weight of the NAD binding domain.

The exact identification of the splitting site of the enzyme by papain to form the 44-kDa fragment (the DNA binding domain) and the 72-kDa fragment was difficult because peptide analysis revealed microheterogeneity of the N-terminal sequence of the 72-kDa fragment. It was noted, however, that three similar amino acid sequences (Thr-Ser-Ala-Ser-Val-Ala, residues 361–366, Ser-Thr-Ala-Ser-Ala-Pro, residues 371–376, and Ser-Ser-Ala-Ser-Ala-Asp, residues 381–386) were coded in the localized region near the N terminus, at which cleavage of peptide bond resulted in the formation of the 44- and the 72-kDa fragments (see Fig. 3). In fact, the N-terminal sequence of Ala-Ser as well as that of Ala-Pro was suggested by peptide analysis of the 72-kDa fragment as described before. Therefore, we concluded that papain cleaved the peptide bond between residues 372–373 as well as those between residues 362–363, residues 374–375, and residues 382–383, resulting in the formation of microheterogeneous fragments of 72 kDa. Based on this conclusion in addition to our earlier observations (7, 8, 11), it is reasonable to consider that the

TABLE III

Amino acid composition of poly(ADP-ribose) synthetase

Residue	Domain						Native Enzyme			
	DNA binding		Automodification		NAD binding		Calculated			Determined
	Residues[a]	%[b]	Residues[a]	%[b]	Residues[a]	%[b]	Residues[a]	%[b]		%[c]
Lys	54	14.52	24	15.79	49	10.02	127	12.54		13.34
His	6	1.61	1	0.66	13	2.66	20	1.97		2.06
Arg	16	4.30	4	2.63	14	2.86	34	3.36		3.43
Asp	25	6.72	4	2.63	33	6.75	62	6.12	9.67	9.94
Asn	7	1.88	7	4.61	22	4.50	36	3.55		
Thr	14	3.76	5	3.29	22	4.50	41	4.05		4.48
Ser	30	8.06	17	11.18	38	7.77	85	8.39		8.46
Glu	30	8.06	15	9.87	30	6.13	75	7.40	10.76	10.96
Gln	14	3.76	3	1.97	17	3.48	34	3.36		
Pro	18	4.84	5	3.29	21	4.29	44	4.34		4.69
Gly	25	6.72	9	5.92	36	7.36	70	6.91		7.43
Ala	22	5.91	17	11.18	26	5.32	65	6.42		6.94
Val	25	6.72	12	7.89	31	6.34	68	6.71		6.22
Met	9	2.42	5	3.29	11	2.25	25	2.47		2.14
Ile	12	3.23	6	3.95	30	6.13	48	4.74		4.61
Leu	25	6.72	14	9.21	51	10.43	90	8.88		9.09
Tyr	9	2.42	0	0.00	23	4.70	32	3.16		3.22
Phe	13	3.49	2	1.32	15	3.07	30	2.96		2.99
Cys	11	2.96	1	0.66	2	0.41	14	1.38		ND[d]
Trp	7	1.88	1	0.66	5	1.02	13	1.28		ND[d]
M_r	(372)	42,018	(152)	16,304	(489)	54,881	(1,013)	113,203		116,000

[a] Residue numbers are calculated from the data in Fig. 3.

[b] Mol/mol of all amino acid residues in each domain or native enzyme.

[c] Data are taken from the paper by Ushiro *et al.* (8).

[d] Not determined.

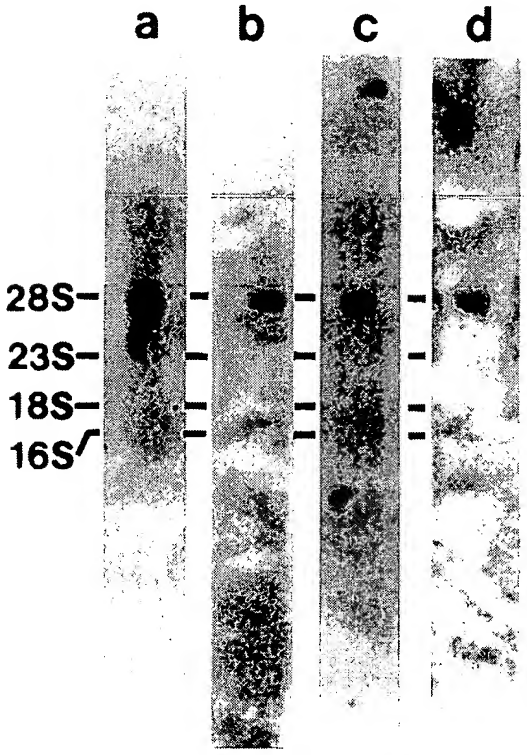

FIG. 4. **Autoradiograms of blot hybridization analysis of human placental poly(A)$^+$ RNA with cDNA probes.** *Bam*HI(2358)-*Bam*HI(2609) 251-bp[2] fragment from pPARS1 (*a*), *Rsa*I(1775)-*Alu*I(1958) 183-bp fragment from pPARS21 (*b*), *Hin*dIII(697)-*Eco*RI(1012) 315-bp fragment from pPARS31 (*c*), and *Rsa*I(153)-*Rsa*I(368) 215-bp fragment from pPARS41(*d*) were used.

DNA binding domain consists of 372 residues, starting from the N-terminal methionine to residue 372 with a calculated molecular weight of 42,018 and that the automodification domain consists of 152 residues starting from residue 373 to residue 524 with a calculated molecular weight of 16,304.[3] This interpretation gives a calculated value of 71,185 for the molecular weight of the 72-kDa fragment.

On the basis of the above consideration, the amino acid composition predicted for each domain of human poly(ADP-ribose) synthetase is presented in Table III. It is noted that the net charge of the NAD binding domain is neutral but the DNA binding domain is considerably rich in lysine residue and quite basic, whereas the automodification domain is relatively basic. Fig. 5 shows a hydrophilicity profile for the deduced amino acid sequence of the native enzyme. As observed in this figure, the enzyme protein appears to be relatively hydrophilic. From the predicted secondary structure, it is estimated that the NAD binding domain appears to be somewhat rich in β-sheet in relation to other domains.

Homology with Other Proteins—A computer search of the protein sequence bank (28, 29, 33) did not reveal any particular protein with a striking overall homology to human poly(ADP-ribose) synthetase. Nevertheless, a separate comparison of the localized sequence of each enzyme domain with those of other proteins indicated that poly(ADP-ribose) synthetase has partial sequence homology with some other proteins. Of particular interest is the fact that the DNA binding

[2] The abbreviations used are: bp, base pair; kb, kilobase; HPLC, high performance liquid chromatography; PTH, phenylthiohydantoin; SDS, sodium dodecyl sulfate; TPCK, L-1-tosyl-amido-2-phenylethyl chloromethyl ketone.

[3] The molecular weight of the automodification domain is somewhat low as compared with that determined experimentally (8). This may be due to the endogenous mono(ADP-ribose) or oligo(ADP-ribose) attached to the domain although direct evidence along this line is lacking.

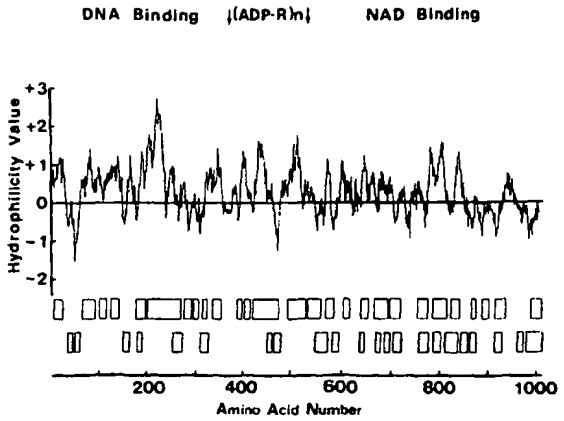

FIG. 5. **Hydrophilicity profile and predicted secondary structure of human poly(ADP-ribose) synthetase.** The averaged hydrophilicity value of an undecapeptide composed of amino acid residues $i - 5$ to $i + 5$ has been plotted against i, where i represents amino acid number. The hydrophilicity values of individual amino acids have been taken from the data of Hopp and Woods (31). The position of the predicted structure of α-helix (*open boxes*) or β-sheet (*dotted boxes*) that has a length of 10 or more residues is shown (32). *Arrows* indicate cleavage sites by papain and by α-chymotrypsin to form three separable enzyme domains.

domain contains the homologous sequence with those of oncogene products such as c-*fos* and v-*fos* (Fig. 6). Also, this domain involves a homologous repeat in the sequence (Fig. 6) and contains a unique sequence element (Fig. 7) which resembles the essential peptide sequences for nuclear location of SV40 and polyoma virus large T antigens (37–39). It is further noted that the NAD binding domain has a sequence similar to the consensus sequence for the binding of adenine nucleotide (Fig. 7) as are observed with various ATPases and adenylate kinase (40).

DISCUSSION

Using the clones isolated, we have determined most of the nucleotide sequence of the cDNA for human poly(ADP-ribose) synthetase. A striking feature of the 5′-untranslated region, although only partially determined, is the G+C richness (71.6%) in relation to the content in the coding region (51.2%). The 3′-untranslated region contains the conical polyadenylation signal 17 bases upstream from the poly(dA) tract. From the coding sequence, we predict that the enzyme consists of 1,013 amino acid residues with a molecular weight of 113,203. This value is very close to that established experimentally (8). The amino acid composition predicted from the cDNA sequence also coincides well with that previously determined (8). Furthermore, the coding sequence involves all of the amino acid sequences of 14 different peptides as obtained by proteolytic digestion of the native enzyme.

Comparison of the total amino acid sequence as deduced from the cDNA with the partial amino acid sequences at the N termini of the 54- and the 72-kDa fragments allowed us to assign the locations of the functionally different three domains of poly(ADP-ribose) synthetase. Based on the amino acid composition predicted for each domain (Table III), the DNA binding domain is rich in lysine residue and quite basic (18.8% basic amino acid residues, the net charge +15). Thus, this domain appears to easily bind to DNA by ionic interaction. In fact, the DNA binding domain from other species is known to be basic as judged by direct amino acid analysis (11). The automodification domain is also rich in lysine residue and basic (the net charge +7). In contrast, the NAD binding domain contains less lysine and arginine residues, and the net charge of this domain is neutral. It is noted that the relative content of glutamic acid residue in the automodification domain, to which poly(ADP-ribose) is attached, is the highest among the three domains of the enzyme protein.

Homology of the total amino acid sequence of poly(ADP-ribose) synthetase with those of other proteins is not so striking. Nevertheless, the partial amino acid sequence of the enzyme somewhat resembles those of several other proteins. Of particular interest is the observation that the DNA binding domain has a homologous repeat in the sequence and exhibits a homology with localized regions of transforming proteins such as c-*fos* and v-*fos* (Fig. 6). Furthermore, this domain contains a unique sequence element similar to those of SV40 and polyoma virus large T antigens which are required for their nuclear localization. Therefore, these facts suggest the possibility that the physiological function of poly(ADP-ribose) synthetase lies in its ability to bind to DNA and to control transformation of living cells like the cases of those oncogene products. Analysis of genomic DNA in normal and transformed eukaryotic cells using cDNA probes isolated in the present study will clarify the above problems.

Acknowledgments—We are grateful to Dr. H. Okayama at National Institutes of Health for providing us with human fibroblast cDNA library, Dr. M. Kanehisa in the Laboratory of Molecular Design of Physiological Functions, Institute for Chemical Research, Kyoto University for his help in computer analysis, and Dr. Y. Ike at Mitsui Toatsu Chemicals, Inc. for the aid in preparing 41-residue-long deoxyribonucleotides. Thanks are also due to S. Yamamoto for preparing competent HB101, S. Shizuta for the aid in screening the data by computer analysis, and M. Isobe and K. Mizuta for secretarial assistance.

REFERENCES

1. Hayaishi, O., and Ueda, K. (1977) *Annu. Rev. Biochem.* **46**, 95–116
2. Ueda, K., and Hayaishi, O. (1985) *Annu. Rev. Biochem.* **54**, 73–100
3. Shizuta, Y., Ito, S., Shizuta, H., and Hayaishi, O. (1978) *Seikagaku* **50**, 919
4. Shizuta, Y., Ito, S., Nakata, K., and Hayaishi, O. (1980) *Methods Enzymol.* **66**, 159–165
5. Ito, S., Shizuta, Y., and Hayaishi, O. (1979) *J. Biol. Chem.* **254**, 3647–3651
6. Agemori, M., Kagamiyama, H., Nishikimi, M., and Shizuta, Y. (1982) *Arch. Biochem. Biophys.* **215**, 621–627
7. Shizuta, Y., Kameshita, I., Agemori, M., Ushiro, H., Taniguchi, T., Ohtsuki, M., Sekimizu, K., and Natori, S. (1985) in *ADP-ribosylation Reactions* (Althaus, F., Hilz, H., and Shall, S., eds.) pp. 52–59 Springer-Verlag, Heidelberg, F. R. G.
8. Ushiro, H., Yokoyama, Y., and Shizuta, Y. (1987) *J. Biol. Chem.* **262**, 2352–2357
9. Nishikimi, M., Ogasawara, K., Kameshita, I., Taniguchi, T., and Shizuta, Y. (1982) *J. Biol. Chem.* **257**, 6102–6105
10. Kameshita, I., Matsuda, Z., Taniguchi, T., and Shizuta, Y. (1984) *J. Biol. Chem.* **259**, 4770–4776
11. Kameshita, I., Matsuda, M., Nishikimi, M., Ushiro, H., and Shizuta, Y. (1986) *J. Biol. Chem.* **261**, 3863–3868
12. Shizuta, Y., Kameshita, I., Ushiro, H., Matsuda, M., Suzuki, S., Mitsuuchi, Y., Yokoyama, Y., and Kurosaki, T. (1986) *Adv. Enzyme Regul.* **25**, 377–384
13. Okayama, H., and Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–289
14. Wahl, G. M., Padgett, R. A., and Stark, G. R. (1979) *J. Biol. Chem.* **254**, 8679–8689
15. Laemmli, U. K. (1970) *Nature* **227**, 680–685
16. Hunkapillar, M. W., Lujan, E., Ostrander, F., and Hood, L. E. (1983) *Methods Enzymol.* **91**, 227–236
17. Konigsberg, W. H., and Henderson, L. (1983) *Methods Enzymol.* **91**, 254–259
18. Hanahan, D., and Meselson, M. (1980) *Gene (Amst.)* **10**, 63–67
19. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
20. Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69**, 1408–1412
21. Rigby, P.W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251
22. Ito, H., Ike, Y., Ikuta, S., and Itakura, K. (1982) *Nucleic Acids Res.* **10**, 1755–1769
23. Grantham, R., Gautier, C., Gouy, M., Jacobzon, M., and Mercier, R. (1981) *Nucleic Acids Res.* **9**, 43–74
24. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
25. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
26. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 5201–5205
27. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
28. Kanehisa, M. (1982) *Nucleic Acids Res.* **10**, 183–196
29. Kanehisa, M. (1986) *IDEAS User Manual*, Program Resources, Inc., Fredrick, MD
30. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872

31. Hopp, T. P., and Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 3824-3828
32. Chou, P. Y., and Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47**, 251-276
33. Locht, C., and Keith, J. M. (1986) *Science* **232**, 1258-1264
34. Berg, J. M. (1986) *Science* **232**, 485-487
35. Berg, J. M. (1986) *Nature* **319**, 264-265
36. Zahradka, P., and Ebisuzaki, K. (1984) *Eur. J. Biochem.* **142**, 503-509
37. Kalderon, D., Richardson, W. D., Markham, A. F., and Smith, A. E. (1984) *Nature* **311**, 33-38
38. Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984) *Cell* **39**, 499-509
39. Richardson, W. D., Roberts, B. L., and Smith, A. E. (1986) *Cell* **44**, 77-85
40. Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) *EMBO J.* **8**, 945-951

Supplementary Material to

Primary Structure of Human Poly(ADP-ribose) Synthetase as Deduced from cDNA Sequence

Tomohiro Kurosaki, Hiroshi Ushiro, Yasuhiro Mitsuuchi, Shigetaka Suzuki, Michiko Matsuda, Yoshiko Matsuda, Nobuhiko Katunuma, Kenji Kangawa, Hisayuki Matsuo, Tadaaki Hirose, Seiichi Inayama and Yutaka Shizuta

EXPERIMENTAL PROCEDURES

Materials---A cDNA library constructed in pcD (13) using poly(A)$^+$RNA from cultured human fibroblasts (nontransformed) was kindly provided by Dr. H. Okayama at National Institutes of Health. For transformation performed by the procedure of Wahl et al. (14), Escherichia coli HB101 was used.

Commercial sources of materials used were as follows: [α-^{32}P]dCTP (3,000 Ci/mmol) and [γ-^{32}P]ATP (5,000 Ci/mmol) from Amersham Corp.; oligo(dT)-cellulose (Type 7) from P-L Biochemicals; PstI-cut and oligo(dG)-tailed pBR322 from Bethesda Research Laboratories; E. coli DNA polymerase I, the Klenow fragment of E. coli DNA polymerase, T4 DNA ligase and T4 polynucleotide kinase from Toyobo Co.; E. coli DNA ligase, E. coli ribonuclease H, S1 nuclease, terminal deoxyribonucleotidyltransferase, bacterial alkaline phosphatase, T4 DNA polymerase and Ball nuclease from Takara Shuzo Co.; reverse transcriptase of avian myeloblastosis virus from Seikagaku Kogyo Co.; restriction endonucleases from Takara Shuzo Co. and Toyobo Co.; TPCK-trypsin from Worthington; α-chymotrypsin from Sigma; bacto tryptone, yeast extract and casamino acids from Difco. All other chemicals were obtained from Nakarai Chemicals (Kyoto) and Wako Pure Chemicals Industries (Osaka) and of analytical grade.

Amino Acid Sequence Analysis---Poly(ADP-ribose) synthetase was purified to homogeneity from fresh human placenta and subjected to limited digestion with papain or α-chymotrypsin as described previously (8). The proteolytic digests were separated by DNA-cellulose chromatography, followed by Bio-Gel P-60 chromatography (8, 11) or SDS-polyacrylamide gel electrophoresis (15). Electroelution was done using the methods of Hunkapiller et al. (16). SDS in each band electroeluted from the gels was removed by extraction using the method of Konigsberg and Henderson (17). Tryptic digestion was performed using 380 ug of poly(ADP-ribose) synthetase and 3 ug of TPCK-trypsin in 50 mM Tris-Cl (pH 8.0) containing 2 mM CaCl, in a final volume of 940 ul. After incubation at 37°C for 15 h, 3 ug of TPCK-trypsin was again added and the incubation was continued for additional 12 h. The tryptic digest was directly loaded onto a column of Chomcosorb 3uC$_{18}$- H(8 x 75 mm; Chemco) equilibrated with solvent A (H$_2$O:acetonitrile:10% trifluoroacetic acid = 90 : 10 : 1). A linear gradient of 10-60% acetonitrile was applied using solvent B (H$_2$O: acetonitrile:10% trifluoroacetic acid = 40 : 60 : 1) over 80 min at a flow rate of 2 ml/min; Hitachi 635 HPLC system with a solvent program was employed. The eluate absorbance at 210 nm and at 280 nm was simultaneously recorded with a Hitachi 41 UV monitor. Twelve different peptides out of 72 absorbance peaks thus obtained were collected and subjected to sequence analysis with a gas-phase sequencer, model 470A, Applied Biosystems. PTH amino acids were analysed by HPLC at 50°C using a column of Ultrasphere ODS(4.6 x 250 mm, Altex).

Cloning Procedures---Table 1 represents a list of cDNA libraries and screening probes employed for selecting the clones containing cDNA inserts for human poly(ADP-ribose) synthetase. Specific oligodeoxyribonucleotides synthesized for preparing primer primed cDNA libraries and the clones selected after screening are also presented in the table.

For first screening of the cDNA library constructed in pcD (13), oligodeoxyribonucleotide probes were radiolabelled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Replica filters were manipulated by the method of Hanahan and Meselson (18) and baked at 80°C for 2 h. They were washed at 60°C for 1 h with 3 x SSC containing 0.1% SDS and then treated at 60°C for 3 h with 50 ug/ml of sonicated calf thymus denatured DNA, 3 x SSC and 10 x Denhardt's solution. Hybridization with two sets of 41 mer (5 x 10^5 cpm/ml) was performed at 55°C for 36 h in 4 x SSC, 10 x Denhardt's solution, 150 ug/ml of sonicated calf thymus denatured DNA and 0.05% sodium pyrophosphate. Filters were washed at 55°C for 15 min with 4 x SSC, next at 55°C for 10 min with 3 x SSC and further at room temperature for 2 min with 4 x SSC. Hybridization with two other sets of 17 mer (5 x 10^4 cpm/ml) was performed at 38°C for 36 h in 4 x SSC, 10 x Denhardt's solution, 150 ug/ml of sonicated calf thymus denatured DNA and 0.05% sodium pyrophosphate. In this case, filters were washed three times at 36°C for 15 min each with 4 x SSC.

Oligo(dT) cDNA library from human placental mRNA was prepared as follows. Total RNA was extracted from fresh human placenta by the guanidine thiocyanate method of Chirgwin et al. (19). Poly(A)$^+$RNA was enriched by oligo(dT)-cellulose chromatography (20). Twenty micrograms of poly(A)$^+$RNA was converted to cDNA in the presence of 50 mM Tris-Cl (pH 8.3), 1 mM each of the deoxyribonucleotide triphosphates, 50 mM KCl, 8 mM MgCl$_2$, 10 mM dithiothreitol, 200 ug/ml of oligo(dT)$_{12-18}$ and 2 units/ul of avian reverse transcriptase in a final volume of 100 ul. Incubation was performed at 42°C for 90 min. The mixture was extracted with phenol-chloroform and the cDNA formed was precipitated with 0.3 M sodium acetate (pH 5.5) and 2.5 volumes of ethanol. The precipitate was collected by centrifugation. The pellet was lyophilized and then dissolved in 150 ul of double distilled water followed by the addition of 30 ul of 1.8 N NaOH containing 50 mM EDTA. The mixture was heated at 70°C for 30 min and then neutralized by the addition of 27 ul of 2 N HCl and 6 ul of 1 M Tris-Cl (pH 7.5). Second strand synthesis was performed in the presence of 100 mM Hepes (pH 6.9), 0.5 mM each of the deoxyribonucleoside triphosphates, 70 mM KCl, 5 mM MgCl$_2$, 15 mM 2-mercaptoethanol and 1 unit/ul of the Klenow fragment of E. coli DNA polymerase in a final volume of 100 ul. Incubation was carried out at 15°C for 10 h. After extraction with phenol-chloroform and precipitation by ethanol, the cDNA was treated at 37°C for 30 min with S1 nuclease at a final concentration of 2 units/ul in 30 mM sodium acetate (pH 4.4), 250 mM NaCl, 1 mM ZnSO$_4$ in a final volume of 250 ul. The mixture was again extracted with phenol-chloroform and precipitated with 2 M ammonium acetate and 2 volumes of ethanol. Oligo(dC) tailing of the cDNA was carried out at 37°C for 15 min in 25 mM Tris-Cl (pH 6.9), 200 mM sodium cacodylate, 0.1 mM dCTP, 2 mM dithiothreitol, 1 mM CoCl$_2$ and 0.7 unit/ul of terminal deoxyribonucleotidyltransferase in a final volume of 100ul. The double stranded cDNA thus prepared was collected and then annealed into the PstI site of oligo(dG)-tailed pBR322. For screening the clones, the restriction fragment as a DNA probe (Table 1) was radiolabelled by nick translation with [α-^{32}P]dCTP (21). Hybridization was performed at 60°C for 18 h with 5 x 10^5 cpm/ml of the DNA probe in 3 x SSC, 10 x Denhardt's solution and 150 ug/ml of sonicated calf thymus denatured DNA. The filters were washed at 60°C for 15 min with 3 x SSC containing 0.1% SDS, then at 60°C for 15 min with 1 x SSC containing 0.1% SDS and finally at room temperature for 1 min with 0.3 x SSC.

For preparing specific oligodeoxyribonucleotide primed cDNA libraries, 100 ug of poly(A)$^+$RNA was mixed with 500 pmol of a corresponding oligodeoxyribonucleotide (see Table 1) in 0.4 M KCl and the solution was incubated at 30°C for 60 min. Primer extension was carried out at 42°C for 90 min in 50 mM Tris-Cl (pH 8.3), 1 mM each of the deoxyribonucleotide triphosphates, 50 mM KCl, 8 mM MgCl$_2$, 10 mM dithiothreitol and 1,200 units/ml of avian reverse transcriptase in a final volume of 1 ml. After extraction with phenol-chloroform and precipitation with 2 M ammonium acetate and 2 volumes of ethanol, second strand synthesis was performed at 15°C for 3 h in 20 mM Tris-Cl (pH 7.5), 0.4 mM each of the deoxyribonucleotide triphosphates, 100 mM KCl, 10 mM ammonium sulfate, 4 mM MgCl$_2$, 225 uM β-NAD, 50 ug/ml bovine serum albumin, 600 units/ml of E. coli DNA polymerase I, 90 units/ml of E. coli ribonuclease H and 90 units/ml of E. coli DNA ligase in a final volume of 1 ml. The mixture was extracted with phenol-chloroform and subjected to precipitation with 2 M ammonium acetate and 2 volumes of ethanol. The cDNA was then treated at 37°C for 30 min with 0.3 unit/ul of T4 DNA polymerase in 33 mM Tris-acetate (pH 6.9), 150 uM each of the deoxyribonucleotide triphosphates, 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM dithiothreitol and 100 ug/ml of bovine serum albumin in a final volume of 70 ul. Oligo(dC) tailing and cDNA insertion into pBR322 were performed by the same procedure as described above. Hybridization with other corresponding DNA probes and filter washing were performed at 60°C using the same cocktails as those employed for screening the oligo(dT) cDNA library. All of the cloning procedures were conducted in accordance with the guideline for research involving recombinant DNA molecules issued by the Ministry of Education, Science and Culture of Japan.

Oligodeoxyribonucleotide Synthesis---The 17 residue and the 41 residue long deoxyribonucleotides were synthesized by the modified triester method (22). They were purified by HPLC and subjected to extensive dialysis prior to use. The base sequences of 41 mers were designed on the basis of the published data for codon usage frequencies (23).

DNA Sequence Analysis---The dideoxy sequencing method (24) using [α-^{32}P]dCTP was employed. Suitable restriction fragments or fragments treated with Bal31 were either subcloned into the M13mp18,19 or the pUC19 vectors (25).

Northern Blot Analysis---The procedure of Thomas (26) was used. Briefly, twenty micrograms of poly(A)$^+$RNA was denatured at 50°C for 1 h in 1 M glyoxal and 50% dimethylsulfoxide. The RNA sample was electrophoresed for 2 h 30 min on a 1.0% agarose gel at a constant voltage of 100. The RNA was transferred from the gel to a nitrocellulose sheet equilibrated with 20 x SSC. After baking at 80°C for 2 h, the sheet was incubated at 42°C for 40 h in the hybridization cocktail containing nick-translated restriction fragment (4 x 10^8 cpm/ml, specific activity 1~3 x 10^8 cpm/ug), 50% formamide, 5 x SSC, 50 mM sodium phosphate buffer (pH 6.5), 250 ug/ml of sonicated salmon sperm denatured DNA, 0.02% bovine serum albumin, 0.02% Ficoll and 0.02% polyvinylpyrrolidone. The sheet was washed four times at room temperature for 5 min with 2 x SSC containing 0.1% SDS and further washed twice at 50°C for 15 min with 0.1 x SSC containing 0.1% SDS. The RNA band hybridized was visualized by exposing the sheet to a New RXO-H film (Fuji) at -70°C for 48 h.

Southern Blot Analysis---Positive clones after screening with oligodeoxyribonucleotide probes were subjected to DNA blot analysis (27). One microgram of DNA, digested with HincII or BamHI was electrophoresed on a 1.5% agarose gel at a constant voltage of 100. The gel was soaked twice at room temperature for 20 min in 0.2 N NaOH and 0.6 M NaCl, and then incubated twice at room temperature for 20 min in 0.2 M Tris-Cl (pH 7.4) containing 0.6 M NaCl. The DNA sample was transferred to a nitrocellulose sheet equilibrated with 6 x SSC. After baked at 80°C for 2 h, the sheet was washed at 60°C for 30 min with 3 x SSC containing 0.1% SDS and again washed at 60°C for 1 h with 3 x SSC, 0.1% SDS and 10 x Denhardt's solution. It was further incubated at 60°C for 1 h in 6 x SSC, 10 x Denhardt's solution, 0.1% SDS, 0.1 mg/ml of sonicated salmon sperm denatured DNA and 0.05% sodium pyrophosphate. The sheet was hybridized at 55°C for 15 h with 2 x 10^5 cpm/ml of the ^{32}P-labelled 41 mer probe in 6 x SSC, 10 x Denhardt's solution, 0.1% SDS, 0.1 mg/ml of sonicated salmon sperm denatured DNA and 0.05% sodium pyrophosphate. The sheet was washed at 55°C for 15 min with 6 x SSC, 0.1% SDS and 0.2% sodium pyrophosphate. The second washing was performed under the same conditions except that 0.3 x SSC was used in place of 6 x SSC. Autoradiography was performed at -70°C for 50 h using a New RXO-H film (Fuji).

Computer Analysis---Data of the nucleotide and the deduced amino acid sequences were analysed by the program of "Integrated Database and Extended Analysis System for Nucleic Acids and Proteins"(IDEAS) according to the method of Kanehisa (28, 29).

RESULTS

Partial Amino Acid Sequence---The initial approach used to clone cDNA sequences for poly(ADP-ribose) synthetase was to determine the partial amino acid sequences of the enzyme protein which provided the data for preparing oligodeoxyribonucleotide probes for isolating cDNA clones. Limited proteolysis of the human placental enzyme with papain yields two fragments of M_r= 44K and 72K, the former corresponding to the DNA binding domain and the latter containing both domains for automodification and for binding of NAD (8, 11). Partial digestion with α-chymotrypsin results in the formation of two fragments of M_r= 54K and 62K, the former corresponding to the NAD binding domain and the latter containing both domains for binding of DNA and for automodification (8, 11). Since each proteolytic fragment of human origin migrated as a broad band or a doublet on SDS-polyacrylamide gel electrophoresis (8), it was further purified as described under "Experimental Procedures" and subjected to amino acid sequence analysis by automated Edman degradation.

The initial analysis of the 44K fragment (the DNA binding domain) as well as the native enzyme revealed no N-terminal sequence; no significant amount of any particular PTH amino acid was recovered during the cycles. This finding is consistent with earlier observations with the enzyme from other species (5, 7) that the N-terminus of poly(ADP-ribose) synthetase is blocked and that the DNA binding domain is located in the N-terminal side of the native enzyme (11). The analysis of the 72K fragment was also unsuccessful due to microheterogeneity of the sample prepared although a sequence of Ala-Ser as well as that of Ala-Pro was suggested.

We then analysed the 54K fragment (the NAD binding domain) and the 40K fragment, the latter derived from the 54K fragment upon prolonged digestion with α-chymotrypsin. Sequencing of the 54K fragment revealed 12 amino acid residues of the N-terminal sequence (Fig. 1A). Similarly, in the 40K fragment identifications were made out to 36 residues with three positions not definitely determined; in cycle 25, 33, and 35 both PTH proline and PTH glutamate were detected (Fig. 1B). The fact that the N-terminal sequence of the 54K fragment is completely different from that of the 40K fragment indicates that the 40K fragment lacks the N-terminal portion of the 54K fragment. In order to obtain some more information of the partial amino acid sequences, poly(ADP-ribose) synthetase was digested with TPCK-trypsin and some of the peptides isolated by the HPLC column were subjected to amino acid sequence analysis as described under "Experimental Procedures". Table II summarizes the results of such experiments. As shown in the table, 12 other amino acid sequences were determined.

Isolation of cDNA Clones---Based on the amino acid sequence of the N-terminus of the 40K fragment, two sets of 41 residue long deoxyribonucleotides were chemically synthesized as described under "Experimental Procedures". Two other sets of 17 residue long mixed deoxyribonucleo-

tides which represented all possible cDNA sequences corresponding to the partial amino acid sequence determined were also prepared (see Table I). Using one set of the 41 residue long deoxyribonucleotides, we first screened the human fibroblasts cDNA library (a total of ~1 x 10^6 transformants) and obtained 28 candidate clones. These candidate clones were further screened with three other sets of oligodeoxyribonucleotide probes and only one positive clone was obtained as judged by blot hybridization analysis (Table I). The nucleotide sequence of the cDNA insert (1.9kb) of this clone (pPARS-F) contained an open reading frame that included the amino acid sequence of the N-terminus of the 40K fragment as well as those of five different tryptic peptides (Table II).

We then screened the oligo(dT) cDNA library from human placenta using the Sau3AI(2866)-EcoRI(3083) fragment of clone pPARS-F as a hybridization probe. Two positive clones were obtained after screening ~3 x 10^5 transformants. The nucleotide sequence analysis of one clone (pPARS1) containing a longer cDNA insert revealed that the sequence of the cDNA insert was exactly identical with the sequence (2298-3221) of clone pPARS-F (Fig. 2). We therefore attempted to elongate the specific oligodeoxyribonucleotide corresponding to the sequence near the 5' end of the cDNA insert of clone pPARS1 by reverse transcription of human placental mRNA and to clone the resulting cDNA transcript into the plasmid pBR322. Screening of this cDNA library by hybridization with the AluI(1988)-AluI(2096) fragment derived from clone pPARS-F led to the isolation of clone pPARS11. The nucleotide sequence of the cDNA insert of clone pPARS11 was identical with the sequence (1963-2314) of clone pPARS-F (Fig. 2).

In parallel experiments, we prepared another cDNA library using human placental mRNA as a template in order to elongate the 17 residue long mixed deoxyribonucleotides which represented all possible cDNA sequences corresponding to the partial amino acid sequence of the 40K fragment. Screening of the library by hybridization with the Sau3AI(1782)-AluI(1958) fragment of the cDNA from clone pPARS-F led to the isolation of clone pPARS21 which coded for an open reading frame including both amino acid sequences of the N-termini of the 54K fragment and the 40K fragment. Furthermore, this clone also coded for an open reading frame corresponding to the tryptic peptide, T-49-2 (Table II).

To select clones harbouring cDNA sequences for a region further upstream, two other sets of specific primer cDNA libraries were prepared (Table I) and 3 clones designated as pPARS31, pPARS32, and pPARS41 were obtained. The nucleotide sequence analyses in comparison with partial amino acid sequences of tryptic peptides (T-49-1, T-36, T-33-1, T-33-2, and T-64-1 in Table II) derived from the purified enzyme indicate that their cDNA sequences involved all residual coding regions including the translational initiation site of the human placental mRNA for poly(ADP-ribose) synthetase.

TABLE I

cDNA Cloning of Poly(ADP-ribose) Synthetase

Source of mRNA	primer	probe	positive colonies[a]	clone selected
(1)Fibroblast	oligo(dT)primed vector(pCD)	1)41mer 5'-ACCATGGCCTTC TTCATG$^{C}_{T}$^{A}GGGCACGTCGAAG ATCATCTT-3'	(28)/ 10^6	
		2)41mer 5'-GGCACGTCGAAG ATCATCTTGATCAGGTCCTGC ACGGGCTT-3'	1/28	pPARS-F
		3)17mer 5'-ACCAT$^{A}_{G}$GC$\{$TT$\{$ TTCAT-3' C		
		4)17mer 5'-AC$^{A}_{G}$TC$^{A}_{G}$AAT$^{A}_{G}$ATC AT$\{$TT-3' C		
(2)Placenta	oligo(dT)$_{12-18}$	5)Sau3AI(2866)-EcoRI(3083) 217 bp fragment from pPARS-F	2/ 3x10^5	pPARS1
(3)Placenta	5'-GCCACCTCGATG TCCAG-3'	6)AluI(1988)-AluI(2096) 108 bp fragment from pPARS-F	9/ 3x10^5	pPARS11
(4)Placenta	5'-ACCAT$^{A}_{G}$GC$\{$TT$\{$ TTCAT-3' C	7)Sau3AI(1782)-AluI(1958) 176 bp fragment from pPARS-F	21/ 3x10^5	pPARS21
(5)Placenta	5'-CCTGACTTCCCT CTTGG-3'	8)DdeI(1361)-DdeI(1416) 55 bp fragment from pPARS21	19/ 3x10^5	pPARS31
		9)HindIII(697)-EcoRI(1012) 315 bp fragment from pPARS31	5/ 3x10^5	pPARS32
(6)Placenta	5'-CAGATCAGGTCG TTCTG-3'	10)RsaI(525)-HindIII(697) 172 bp fragment from pPARS32	17/ 3x10^5	pPARS41

a/ Number of hybridization positive colonies/number of screened colonies

TABLE II

Amino Acid Sequence of Tryptic Peptides from Poly(ADP-ribose) Synthetase

	Peptide[a]	Amino Acid Sequence[b]	Region(Residue No.)[c]
1.	T-49-1	MAIMVQSPMFDGK	35-47
2.	T-36	HPDVEVDGFSELR	66-78
3.	T-33-1	TLGDFAAEYAK	109-119
4.	T-33-2	EELGFRPEYSASQLK	168-182
5.	T-64-1	VADGMVFGALLPCEECSGQLVF	283-304
6.	T-49-2	LAHILSPWGAEVK	473-485
7.	T-59-1	VFSATLGLVDIVK	551-563
8.	T-62	SILSEVQQAVSQGSSDSQILDLSNR	710-734
9.	T-56	FYTLIPHDFGMK	735-746
10.	T-64-2	VEMLDNLLDIEVAY	761-774
11.	T-59-2	NTHATTHNAYDLEVIDIFK	819-837
12.	T-65	TTPDPSANISLDGVDVPLG	953-971

a/ Detailed data of peptide isolation will be published elsewhere(Matsuda et al., manuscript in preparation).
b/ The one-letter amino acid notation is used in the table.
c/ See residue numbers in Fig 3.

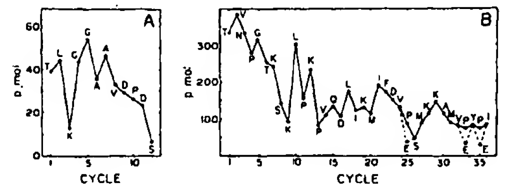

Fig. 1. Sequence analysis of peptides. α-chymotryptic fragments of poly(ADP-ribose) synthetase from human placenta were prepared and subjected to amino acid sequence analysis as described under "Experimental Procedures". The yields of PTH amino acids recovered from the 54K fragment (A) and the 40K fragment (B) at each cycle of Edman degradation are shown; one-letter amino acid notation is used.

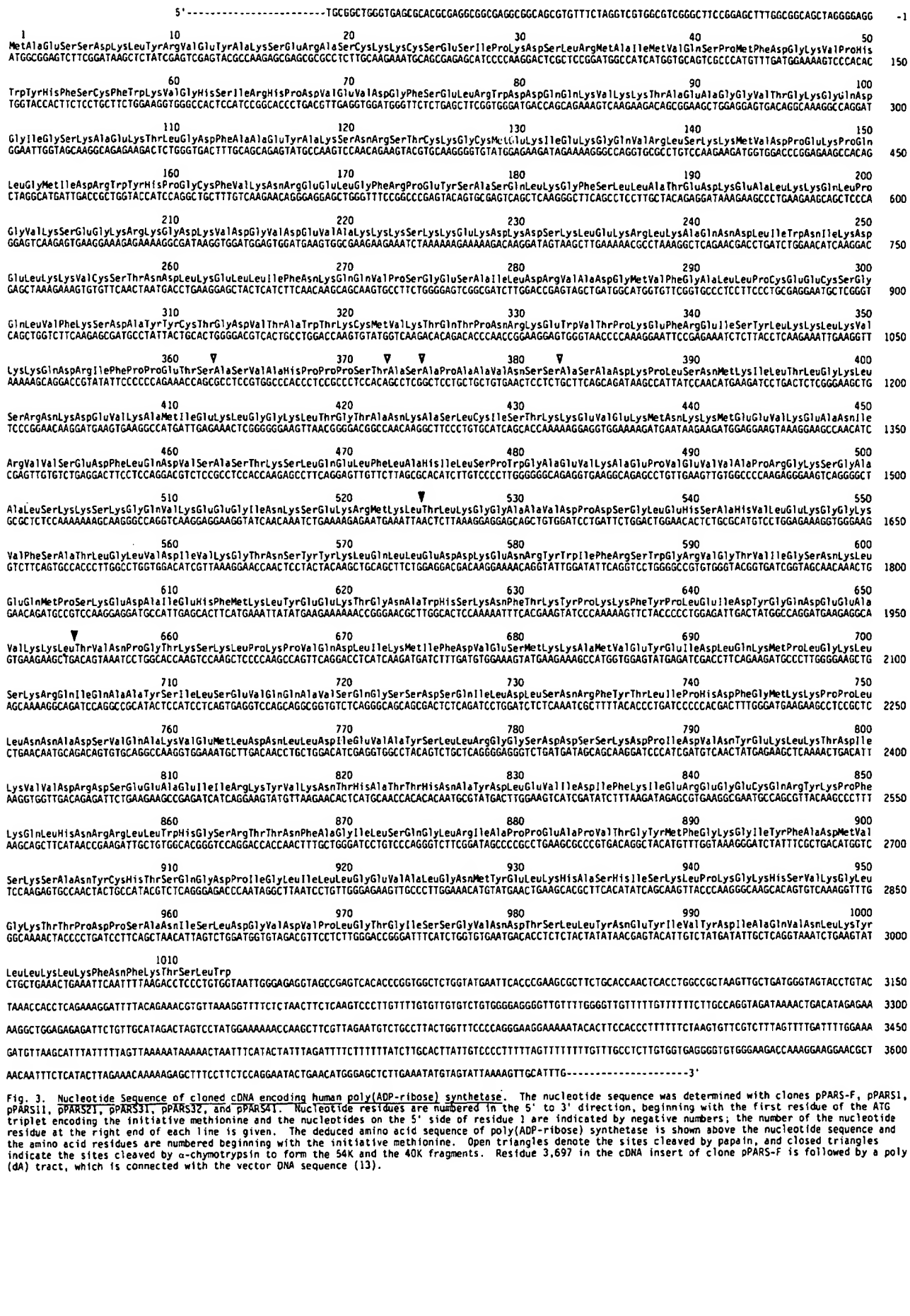

Fig. 3. Nucleotide Sequence of cloned cDNA encoding human poly(ADP-ribose) synthetase. The nucleotide sequence was determined with clones pPARS-F, pPARS1, pPARS11, pPARS21, pPARS31, pPARS32, and pPARS41. Nucleotide residues are numbered in the 5' to 3' direction, beginning with the first residue of the ATG triplet encoding the initiative methionine and the nucleotides on the 5' side of residue 1 are indicated by negative numbers; the number of the nucleotide residue at the right end of each line is given. The deduced amino acid sequence of poly(ADP-ribose) synthetase is shown above the nucleotide sequence and the amino acid residues are numbered beginning with the initiative methionine. Open triangles denote the sites cleaved by papain, and closed triangles indicate the sites cleaved by α-chymotrypsin to form the 54K and the 40K fragments. Residue 3,697 in the cDNA insert of clone pPARS-F is followed by a poly (dA) tract, which is connected with the vector DNA sequence (13).

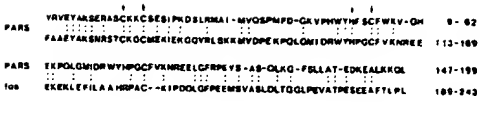

Fig. 6. Homologous repeat of the amino acid sequence of the DNA binding domain of poly(ADP-ribose) synthetase (top) and the comparison of homologous region of the DNA binding domain with the human c-fos protein (bottom). One-letter amino acid notation is used. PARS stands for poly(ADP-ribose) synthetase. Arrows show possible metal binding ligands in the DNA binding domain (34, 35) as poly(ADP-ribose) synthetase is shown to be a zinc metalloenzyme (36).

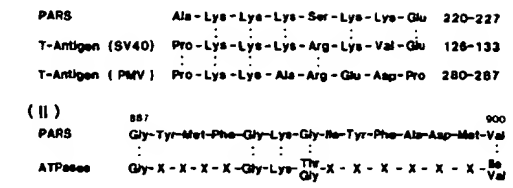

Fig. 7. (I) Comparison of the unique sequence element of the DNA binding domain with the essential peptide sequences for nuclear location of SV40 and polyoma virus large T antigens. (II) Comparison of the sequence of the NAD binding domain with the consensus sequence for the binding of adenine nucleotide (40). In the above figure, PARS stands for poly(ADP-ribose) synthetase, PMV represents polyoma virus and X can be any amino acid residue. Note that two other unique sequence elements (Lys-Lys-Leu-Lys-Val-Lys-Lys and Lys-Arg-Lys-Gly-Asp-Lys-Val) which may specifically recognize the DNA ligand are present in the DNA binding domain (residues 346-352 and residues 207-213 in Fig. 3). Also, note that two other sequences similar to that for the binding of adenine nucleotide are present in the DNA binding domain (residues 92-106 and residues 201-216 in Fig. 3).